

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2002, 00:09:14 ; Search time 8498.8 Seconds  
(without alignments)  
36.667 Million cell updates/sec

Title: US-09-851-670-5

Perfect score: 29  
Sequence: 1 ttgcttgctgctgctgctgcttca 29

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 111874

Minimum DB seq length: 0  
Maximum DB seq length: 60

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estlpl:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hlc:\*  
10: qb\_estl1:\*  
11: qb\_estl2:\*  
12: qb\_hlc:\*  
13: qb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_iny:\*  
17: em\_gss\_pin:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rtd:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17.4	60.0	50	10	AU105368 AU105368
2	15.8	54.5	35	13	BH023779 BC02293-3
3	15.4	53.1	54	13	AZ447232 IM0244L13
4	15.2	52.4	40	13	CNS06VE6 T3 end of
5	15.2	52.4	44	13	AZ505846 IM0346P07
6	15.2	52.4	55	10	AA906735 OK78H04.S
7	15.2	52.4	56	13	AZ930011 479.G1157
8	15.1	51.7	26	13	AZ307056 IM0008G23
9	14.8	51.0	36	13	AZ330732 IM0056D12
10	14.8	51.0	39	13	AZ330739 IM0056F12
11	14.8	51.0	58	13	AZ424808 IM0204A07
12	14.8	51.0	60	10	AA405801 zt57f07.S

13	14.6	50.3	40	13	AZ391073	AZ391073 IM0152124
14	14.6	50.3	58	13	AZ661852	AZ661852 IM0540F07
15	14.4	49.7	41	13	AZ459776	AZ459776 IM0264A19
16	14.4	49.7	52	10	AA894517	AA894517 cF90A05.S
17	14.2	49.0	29	13	AZ827060	AZ827060 IM0103F17
18	14.2	49.0	32	13	AZ627466	AZ627466 IM0469I13
19	14.2	49.0	34	13	AZ626219	AZ626219 IM0466B23
20	14.2	49.0	36	10	AM248008	AM248008 2820129..3
21	14.2	49.0	37	13	AZ820288	AZ820288 2M0092G19
22	14.2	49.0	38	13	AZ819005	AZ819005 2M0089F23
23	14.2	49.0	40	10	AA888211	AA888211 cF86d11.S
24	14.2	49.0	41	13	AZ356235	AZ356235 IM0097D16
25	14.2	49.0	43	10	A1018532	A1018532 cu24g12.x
26	14.2	49.0	45	13	AZ634992	AZ634992 IM0491A08
27	14.2	49.0	46	11	T25667	T25667 EST00536.B4
28	14.2	49.0	50	10	AU103997	AU103997 AU103997
29	14.2	49.0	50	10	BE732496	BE732496 601567696
30	14.2	49.0	50	11	BG538913	BG538913 602568565
31	14.2	49.0	51	11	C00246	C00246 HUMGS000591
32	14.2	49.0	52	10	AM410447	AM410447 FH06C03.Y
33	14.2	49.0	55	11	BG694530	BG694530 NISC_1v03
34	14.2	49.0	60	10	AW058965	AW058965 fe14d12.x
35	14.2	48.3	38	11	H99519	H99519 yx29b12.s1
36	14.2	48.3	43	13	AZ960584	AZ960584 2M0228M07
37	14.2	48.3	50	13	AQ025638	AQ025638 1(2)05488
38	13.8	47.6	52	10	AA626479	AA626479 ab50a01.r
39	13.8	47.6	55	13	AZ333386	AZ333386 IM0062N23
40	13.8	47.6	55	13	AZ658522	AZ658522 IM0535A18
41	13.8	47.6	58	13	AZ537183	AZ537183 AST-2P026
42	13.6	46.9	36	13	AZ853353	AZ853353 2M0156D06
43	13.6	46.9	44	10	A1540308	A1540308 TG34602.X
44	13.6	46.9	43	13	AQ074129	AQ074129 ER(X)1612
45	13.6	46.9	49	13	AZ876952	AZ876952 2M0192E12

#### ALIGNMENTS

RESULT 1	AU105368	50 bp	mRNA	EST	05-APR-2001
LOCUS	AU105368	Sugano Homo sapiens cDNA library	Homo sapiens cDNA clone		
DEFINITION	HEMB0039	mRNA sequence.			
ACCESSION	AU105368				
VERSION	AU105368.1	GI:13554889			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
AUTHORS	Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata,H., Ota,T., Isegaki,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo,K., Suyama,A. and Sugano,S.				
TITLE	fine structural analysis of transcription start sites of human				
JOURNAL	mRNAs using full-length enriched and 5'-end enriched cDNA libraries				
COMMENT	Unpublished (2001)				
CONTACT	Department of Medical Science, University of Tokyo				
DEPARTMENT	4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan				
EMAIL	Email: yszukie@ms.u-tokyo.ac.jp				
FEATURES	Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano,S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).				
SOURCE	Location/Qualifiers				
	1..50				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="HEMB0039"				
	/clone_id="Sugano Homo sapiens cDNA library"				
BASE COUNT	9 a	7 c	11 g	22 t	1 others
ORIGIN					

Query Match 60.0%; Score 17.4; DB 10; Length 50;  
 Best Local Similarity 75.0%; Pred. No. 2.2e+04;  
 Matches 21; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 15 TTGGNATGCGCATTTGTTTCTTTTA 42

RESULT 2  
 LOCUS BH023779 35 bp DNA GSS 09-JUL-2001  
 DEFINITION BG02293-3prime Drosophila melanogaster P[Gri] P element insertion lines Drosophila melanogaster genomic Sequence recovered from 3' end of P element, DNA sequence.

ACCESSION BH023779  
 VERSION BH023779.1 GI:14627238  
 KEYWORDS GSS.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 Lewis, R., Hoskins, R., Liao, G., Morzen, N., Tsang, G., He, Y., Karpen, G., Beilen, H., Rubin, G., and Spradling, A.  
 The Berkeley Drosophila Genome Project Gene Disruption Project  
 Unpublished (2001)  
 Contact: Gerald Rubin  
 Berkeley Drosophila Genome Project  
 University of California, Berkeley  
 LSA Building Berkeley, CA 94720-3200, USA  
 Fax: 5106439947  
 Email: gerry@fruitfly.berkeley.edu  
 Sequence recovery method was inverse PCR.  
 Sequence orientation is forward strand relative to 5' end of P element  
 The P element insertion position is base 1 in the 35 bases. This insertion position refers to the first base of the 8 base target recognition sequence.  
 Class: transposon-tagged.  
 Location/Qualifiers  
 1..35  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_lib="Drosophila melanogaster P[Gri] P element insertion lines"  
 /note="Inverse PCR was performed on Drosophila melanogaster strains each of which contains one or more P[Gri] P-element transposon insertion. The resultant fragment for each strain was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at <http://www.fruitfly.org/about/methods/inverse.pcr.html>."

BASE COUNT 0 a 12 c 9 g 14 t

ORIGIN

Query Match 54.5%; Score 15.8; DB 13; Length 35;  
 Best Local Similarity 74.1%; Pred. No. 7.9e+04;  
 Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 7 TTGCGTTGCGTCTGTTTCTTTTC 33

RESULT 3  
 LOCUS A2447232 54 bp DNA GSS 04-OCT-2000  
 DEFINITION IM02441L3F Mouse 10kb plasmid UNGC1M library Mus musculus genomic

ACCESSION clone UNGC1M02441L3 F, DNA sequence.  
 A2447232  
 VERSION A2447232.1 GI:10599012  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 54)  
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmud, M., Meenen, E., Pedersen, T., Reilly, and Wright, D., Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0244 row: L column: 13  
 Seq primer: CGTTGTAAACGACGCGCAGT  
 Class: plasmid ends  
 High quality sequence stop: 54.  
 Location/Qualifiers  
 1..54  
 /organism="Mus musculus"  
 /strain="C57Bl/6j"  
 /db\_xref="taxon:10090"  
 /clone="UNG1M02441L3"  
 /clone\_lib="Mouse 10kb plasmid UNGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57Bl/6j (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
 (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (911473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 28 a 11 c 4 g 11 t

ORIGIN

Query Match 53.1%; Score 15.4; DB 13; Length 54;  
 Best Local Similarity 76.0%; Pred. No. 1e+05;  
 Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 33 GCTTGTGTGTGTGATGATGTTCA 9

RESULT 4  
 LOCUS CNS06VE6 40 bp DNA GSS 06-JUL-2001  
 DEFINITION CNS06VE6

```

DEFINITION  T3 end of clone AX0A019F02 of library AX0A from strain CBS 7064
ACCESSION   AL417028
VERSION     AL417028.1 GI:12198231
KEYWORDS    GSS.
SOURCE      Pichia farinosa.
ORGANISM    Pichia farinosa
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1 (bases 1 to 40)
AUTHORS     de Montigny,J., Spohner,C., Souciet,J., Tekala,F., Dujon,B.,
            Wincker,P., Artiguenave,F. and Potier,S.
TITLE        Genomic exploration of the hemiascomycetous yeasts: 15. Pichia
            sorbitophila
JOURNAL      FEBS Lett. 487 (1), 87-90 (2000)
MEDLINE     20584725
REFERENCE   2 (bases 1 to 40)
AUTHORS     Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bojotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Leplingle,A., Llorente,B.,
            Malpertuy,A., Neuvéglise,C., Ozler-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekala,F., Toffiano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
TITLE        Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
JOURNAL      FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE     20584711
REFERENCE   3 (bases 1 to 40)
AUTHORS     Genoscope.
TITLE        Direct Submission
JOURNAL      Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.

FEATURES
    source
        1..40
            /organism="Pichia farinosa"
            /strain="CBS 7064"
            /db_xref="taxon:4920"
            /clone="AX0A019F02"
            /clone_1lb="AX0A"
            /note="end : '73'"

BASE COUNT      0 a          9 c          4 g          26 t          1 others
ORIGIN
Query Match      52.4%; Score 15.2; DB 13; Length 40;
Best Local Similarity 71.4%; Pred. No. 1.2e+05;
Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ttggcttgctgctgctgctgctt 28
    ||||| ||| | | | ||||| |||||
Db 3 TTGGCTTCGCTTTGGCTTCCTTTTTC 30

RESULT 5
A2505846 44 bp DNA GSS 05-OCT-2000
LOCUS     IM0346P07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION
ACCESSION A2505846
VERSION   A2505846.1 GI:10687162
KEYWORDS GSS.
SOURCE    house mouse.

```

```

ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 44)
AUTHORS       Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly
            M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
TITLE         Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL       Unpublished (2000)
COMMENT       Contact: Robert B. Weiss
            University of Utah Genome Center
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0346 row: P column: 07
            Seq primer: CACACAGCAACGCTATGACC
            Class: Plasmid ends
            High quality sequence stop: 44.

FEATURES
    source
        1..44
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="UUGC1M0346P07"
            /clone_1lb="Mouse 10kb plasmid UUGC1M library"
            /sex="Male"
            /lab_host="E. Coli strain XL10-Gold, r1-resistant, F-"
            /note="Vector: pMD42ny; Purified genomic DNA from M.
            musculus (C57BL/6J male) was obtained from the Jackson
            Laboratory Mouse DNA Resource
            (http://www.jax.org/resources/documents/dnares/). The DNA
            was hydrodynamically sheared by repeated passage through a
            0.005 inch orifice at constant velocity. The sheared DNA
            was blunt end-repaired with T4 DNA polymerase and T4
            polynucleotide kinase. Adaptor oligonucleotides were
            ligated to the blunt ends in high molar excess. The
            adaptor DNA was purified and size-selected for a 9.5 to
            10.5 kb range using preparative agarose gel
            electrophoresis. Vector DNA was prepared from a derivative
            of pMD42 (g1147321149b/AF129072.1), a copy-number
            inducible derivative of plasmid R1. The vector was ligated
            with adaptors complementary to the insert adaptors and
            purified. The sheared, adaptor mouse DNA was annealed to
            adaptor vector DNA, and transformed into
            chemically-competent E. coli XL10-Gold (Stratagene) cells
            and selected for ampicillin resistance."

BASE COUNT      15 a          12 c          8 g          9 t
ORIGIN
Query Match      52.4%; Score 15.2; DB 13; Length 44;
Best Local Similarity 71.4%; Pred. No. 1.2e+05;
Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 ttggcttgctgctgctgcttcttca 29
    ||||| ||| | | | ||||| |||||
Db 39 TGGGCTTTGGACGCTTATCTTTTCACA 12

RESULT 6
AA906735 55 bp mRNA EST 24-AUG-1998
LOCUS     OK78h04.s1 NCI-CGAP-CC4 Homo sapiens cDNA clone IMAGE:1520119 3'
DEFINITION
ACCESSION AA906735
VERSION   AA906735.1 GI:3042321
KEYWORDS sequence.
SOURCE    similar to TR:Q34192 Q34192 NAH3 DEHYDROGENASE SUBUNIT 5. ; mRNA

```

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 55)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
Emmett-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/RLNL at:  
www.bio.llnl.gov/dbp/image/image.html

Trace considered overall poor quality  
Insert Length: 378 Std Error: 0.00  
Seq primer: -40m3 fwd. ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers  
1. .55  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1520119"  
/clone\_lib="NCI-CGAP\_GC4"  
/tissue\_type="pooled germ cell tumors"  
/lab\_host="DH10B"  
/note="Vector: pF73D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from 3 pooled  
germ cell tumors, and was then primed with a Not I -  
oligo(dT) primer. Double-stranded cDNA was ligated to Eco  
RI adaptors (Pharmacia), digested with Not I and cloned  
into the Not I and Eco RI sites of the modified pF73  
vector. Library is normalized. Library was constructed by  
Bento Soares and M. Fatima Bonaldo."

BASE COUNT 0 a 4 c 11 g 40 t  
ORIGIN

Query Match 52.4%; Score 15.2; DB 10; Length 55;  
Best Local Similarity 71.4%; Pred. No. 1.2e+05;  
Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 ttggcttgctgcgttcgttcgttc 28  
||||| ||||| ||||| ||||| |||||  
Db 26 TTTTGGTTTGTCTTCTGCTTTTTC 53

RESULT 7  
LOCUS A2930011 56 bp DNA GSS 01-APR-2001  
DEFINITION 479.d1157f11.s1 Saccharomyces kluyveri Saccharomyces kluyveri  
genomic clone 479.d1157f11.s1, DNA sequence.  
ACCESSION A2930011  
VERSION A2930011.1 GI:13500919  
KEYWORDS GSS.  
SOURCE Saccharomyces kluyveri.  
ORGANISM Saccharomyces kluyveri.  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
REFERENCE 1 (bases 1 to 56)  
AUTHORS Clifton, P.F., Hillier, L.M., Fulton, L., Graves, T., Miner, T., Gish  
W.R., Waterston, R.H. and Johnston, M.  
TITLE Surveying Saccharomyces genomes to identify functional elements by  
comparative DNA sequence analysis  
JOURNAL Unpublished (2001)  
COMMENT Contact: Johnston M

Department of Genetics  
Washington University Medical School  
Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA  
Tel: 314 362 2735  
Fax: 314 362 7855  
Email: mj@genetics.wustl.edu  
Class: random plasmid subclone.  
location/Qualifiers  
1. .56  
/organism="Saccharomyces kluyveri"  
/strain="NRRL Y-12651 (CBS 3082)"  
/db\_xref="taxon:4934"  
/clone="479.d1157f11.s1"  
/clone\_lib="Saccharomyces kluyveri"  
/note="Random genomic sequence"

BASE COUNT 12 a 11 c 13 g 20 t  
ORIGIN

Query Match 52.4%; Score 15.2; DB 13; Length 56;  
Best Local Similarity 71.4%; Pred. No. 1.2e+05;  
Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 ttggcttgctgcgttcgttcgttc 28  
||||| ||||| ||||| ||||| |||||  
Db 13 TTTGGTTTGTCTGCTGATACCTTC 40

RESULT 8  
LOCUS A2307056 26 bp DNA GSS 29-SEP-2000  
DEFINITION 1M0008G23F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0008G23 F, DNA sequence.  
ACCESSION A2307056  
VERSION A2307056.1 GI:10345677  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 26)  
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.  
and Wright, D. Weis, R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weis  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert length: 10000 Std Error: 0.00  
Plate: 0008 row: G column: 23  
Seq primer: CTTGTAAACGACGCGCACT  
Class: plasmid ends  
High quality sequence stop: 26.  
location/Qualifiers  
1. .26  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0008G23"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: pMD29v; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 0 a 2 c 11 g 13 t  
ORIGIN

Query Match 51.7%; Score 15; DB 13; Length 26;  
Best Local Similarity 78.3%; Pred. No. 1.5e+05;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ttgcttggtgctgctgctgctg 23  
|||||  
Db 4 TCTGGCTTTGGTGGTGTCTG 26

RESULT 9  
A2330732 36 bp DNA GSS 29-SEP-2000  
LOCUS A2330732/c  
DEFINITION 1M0056F12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0056F12 F, DNA sequence.  
ACCESSION A2330732  
VERSION A2330732.1 GI:10392727  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus.

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 36)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamli,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0056 row: D column: 12  
Seq primer: CGTGTAAACGACGCCACGT  
Class: plasmid ends  
High quality sequence stop: 36.

## FEATURES

Source  
1. 36  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0056F12"  
/clone\_1lb="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 24 a 10 c 0 g 2 t  
ORIGIN

Query Match 51.0%; Score 14.8; DB 13; Length 36;  
Best Local Similarity 73.1%; Pred. No. 1.7e+05;  
Matches 19; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 ttgcttggtgctgctgctgctg 27  
|||||  
Db 31 TTGGTTTGTGTTGATTTGTTTT 6

RESULT 10  
A2330739 39 bp DNA GSS 29-SEP-2000  
LOCUS A2330739/c  
DEFINITION 1M0056F12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0056F12 F, DNA sequence.  
ACCESSION A2330739  
VERSION A2330739.1 GI:10392741  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus.

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 39)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamli,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0056 row: F column: 12  
Seq primer: CGTGTAAACGACGCCACGT  
Class: plasmid ends  
High quality sequence stop: 39.

## FEATURES

Source  
1. 39  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0056F12"  
/clone\_1lb="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: pMD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 25 a 12 c 0 g 2 t  
ORIGIN

Query Match 51.0%; Score 14.8; DB 13; Length 39;  
Best Local Similarity 73.1%; Pred. No. 1.7e+05;  
Matches 19; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ttggcttgctgcgtcgtcgttctt 27  
||||| 11 11 1111111  
Db 33 TTGGTTTTGTTTGTGTTTGT 8

RESULT 11  
LOCUS A2424808 58 bp DNA GSS 03-OCT-2000  
DEFINITION IM0204A07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
ACCESSION A2424808  
VERSION A2424808.1 GI:10548821  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scurionath; Muridae; Mus.  
1 (bases 1 to 58)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A., and Wright, D., Weis, R.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0204 row: A column: 07  
Seq primer: CACACAGGAACAGCATGACC  
Class: plasmid ends  
High quality sequence stop: 58.  
Location/Qualifiers  
1. 58  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0204A07"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 7 a 1 c 17 g 33 t  
ORIGIN

Query Match 51.0%; Score 14.8; DB 13; Length 58;  
Best Local Similarity 73.1%; Pred. No. 1.6e+05;  
Matches 19; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ttggcttgctgcgtcgtcgttctt 26  
||||| 11 11 1111111  
Db 30 TTTTGTGTTGTTGTTGTTGTT 55

RESULT 12  
LOCUS AA405801 60 bp mRNA EST 09-NOV-1997  
DEFINITION ZNF767.s1 Soares ovary tumor NBHOT Homo sapiens cDNA clone  
IMAGE:742117 3', mRNA sequence.  
ACCESSION AA405801  
VERSION AA405801.1 GI:2063784  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 60)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, C., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Matra, M., Martin, J., Moore, B., Schellenberg, K., Stepove, M., Tan, F., Theisling, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
WashU-NCI human EST Project  
Unpublished (1997)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.  
Insert Length: 171 Std Error: 0.00  
Seq primer: -41ml3 fwd. ET from Amersham.  
Location/Qualifiers  
1. 60  
/organism="Homo sapiens"  
/db\_xref="GDB:5942423"  
/db\_xref="taxon:9606"  
/clone="IMAGE:742117"  
/clone\_lib="Soares ovary tumor NBHOT"  
/sex="Female"  
/tissue\_type="ovarian tumor"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: ovary; Vector: pTV73D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st

lab host E. coli strain X10-Gold, T1-resistant, F<sup>-</sup>/hcs-Vector (PMD419). Purified genomic DNA from *M. musculus* C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel

```

/sex="Male"
/lab_host="E. coli strain XL10-gold, T1-resistant, F-"  

/note="Vector: pMD22hy. Purified genomic DNA from M.  

musculus C57Bl/6J (male) was obtained from the Jackson  

Laboratory Mouse DNA Resource  

(http://www.jax.org/resources/documents/dnares/). The DNA  

was hydrodynamically sheared by repeated passage through a  

0.005 inch orifice at constant velocity. The sheared DNA  

was blunt end-repaired with T4 DNA polymerase and T4  

polynucleotide kinase. Adaptor oligonucleotides were  

ligated to the blunt ends in high molar excess. The  

adaptor DNA was purified and size-selected for a 9.5 to

```

10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g1147321149b|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 0 a 0 c 21 g 37 t  
ORIGIN

Query Match 50.3%; Score 14.6; DB 13; Length 58;  
Best Local Similarity 81.0%; Pred. No. 1.9e+05;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 7 ttgtgctgcgtcgtctgttt 27  
||||| ||||| ||||| ||||| |||||  
Db 25 TTGTCTTGTCTTGTCTTGT 45

RESULT 15  
A2459776 41 bp DNA GSS 04-OCT-2000  
LOCUS 1M0264A19R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
DEFINITION clone UUGC1M0264A19 R, DNA sequence.  
ACCESSION A2459776  
VERSION A2459776.1 GI:10617901  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 41)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLc, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunne@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0264 row: A column: 19  
Seq primer: CACACAGGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 41.  
Location/Qualifiers

# FEATURES

1. 41  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0264A19"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g1147321149b|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 2 a 3 c 12 g 24 t  
ORIGIN

Query Match 49.7%; Score 14.4; DB 13; Length 41;  
Best Local Similarity 75.0%; Pred. No. 2.3e+05;  
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 ggccttgctgcgtcgtcgttt 27  
|| ||| || ||||| ||| |||  
Db 4 GCTTTTGTGTGATCTTGT 27

Search completed: March 9, 2002, 00:09:17  
Job time: 11033 sec



**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**